



HAL
open science

L'onde sonore : réalités physiques et perception

Alain Ghio

► **To cite this version:**

Alain Ghio. L'onde sonore : réalités physiques et perception. Auzou P.; Rolland V.; Pinto S. ; Ozsancak C. Les dysarthries, Solal, 2007. hal-01616690

HAL Id: hal-01616690

<https://hal.science/hal-01616690>

Submitted on 17 Oct 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

L'onde sonore : réalités physiques et perception

A. Ghio - Ingénieur de recherche

*Laboratoire Parole et Langage, CNRS UMR 6057
Université de Provence
29, Avenue Robert Schuman
13 621 Aix-en-Provence, CEDEX 1.*

Courriel : alain.ghio@lpl.univ-aix.fr

Au sens de la théorie de l'information, la communication parlée est un processus d'émission/réception d'un message dont le support physique est constitué principalement par une onde acoustique. Aussi, nous exposerons dans ce chapitre des notions générales d'acoustique mais aussi des aspects plus spécifiques de la parole.

I. L'onde acoustique

1. Qu'est-ce qu'une onde acoustique ?

Une onde acoustique correspond à un ébranlement de l'air. Sous l'effet d'une excitation mécanique comme, par exemple, une membrane de haut-parleur, des molécules d'air reçoivent une impulsion qui les met en mouvement dans une certaine direction. Elles rencontrent alors d'autres molécules qu'elles poussent devant elles, formant ainsi une zone de compression. L'air étant élastique, la zone comprimée ne tarde pas à se détendre, ce qui entraîne une compression dans la région limitrophe. Le phénomène se propage ainsi de proche en proche comme le renversement d'une rangée de dominos. Si, d'un point de vue macroscopique, l'onde peut se déplacer sur une grande distance, le mouvement des molécules reste très local, de la même façon que la chute de dominos reste très localisée alors que visuellement, l'avalanche peut se propager sur une longueur importante. Par conséquent, un haut-parleur, même très puissant, ne peut pas créer de courant d'air.

Pour se propager, une onde acoustique a besoin d'un support matériel comme, par exemple, l'air où elle se déplace à environ 340 mètres par seconde (m/s), dans l'eau à 1450 m/s, dans les os à 3500 m/s... Cette dernière caractéristique explique la particularité de la perception que l'on a de sa propre parole, le locuteur bénéficiant à la fois d'une conduction osseuse et d'une perception purement aérienne de sa voix. Voilà pourquoi nous percevons très différemment des enregistrements de notre parole par rapport à notre sensation 'on-line', car dans le premier cas, seule l'onde aérienne est perçue.

La production de parole s'apparente plus à un instrument à vent qu'à une simple membrane vibratoire de type haut-parleur. Il s'agit d'un «hachage» de l'air expulsé par les poumons à travers les cordes vocales. Ainsi, en phonation, on observe deux phénomènes simultanés de mouvements d'air : le flux d'air «expiratoire» qui se déplace en colonne et le mouvement local des molécules mises en mouvement localement par l'onde acoustique. On pourrait comparer ce phénomène au mouvement de l'eau dans un fleuve après y avoir jeté une pierre : l'onde créée par la pierre entraîne un mouvement local des molécules d'eau en forme de cercle grandissant autour de l'impact mais la masse d'eau se déplace aussi globalement avec le courant. De ce fait, dans la production de parole, les dispositifs d'enregistrement doivent être adaptés soit à la mesure de l'onde acoustique, soit à la détection de la colonne d'air expiratoire.

Fréquemment, dans les ouvrages d'acoustique, est mentionnée la notion d'onde plane. Il s'agit d'un concept issu de la physique de la propagation des ondes. Une onde plane est une onde dont les fronts d'onde sont des plans infinis, parallèles entre eux et perpendiculaires à une même direction de propagation. Il s'agit d'un cas singulier important, en particulier pour le physicien théoricien. Dans la réalité, une onde plane pure n'existe pas, mais il est possible de s'en approcher dans un domaine limité de l'espace. Il suffit que les fronts d'onde soient suffisamment plans et parallèles dans le volume d'intérêt. Une méthode pour obtenir une approximation d'onde plane partant d'une onde sphérique est de se placer en champ lointain : le rayon de courbure du front d'onde sphérique devient alors très grand. Concrètement, une petite portion d'une onde quelconque peut être considérée comme plane quand la distance à la source dépasse quelques longueurs d'onde et en l'absence de réflexions (Botte et al., 1988).

2. Qu'est-ce qu'un son et comment l'enregistrer ?

Un son correspond à un ébranlement répétitif des molécules d'air, le phénomène de compression et de détente se propageant de proche en proche et de façon répétitive sous l'effet d'une excitation mécanique alternative. Il faut noter que l'ordre de grandeur des pressions acoustiques reste faible : le seuil d'audibilité correspond à une valeur de 0.00002 Pascal (Pa), une conversation soutenue à environ 1 Pa, un avion à réaction à 200 Pa. Ces valeurs apparaissent dérisoires comparées à la pression atmosphérique de l'ordre de 100 000 Pa. De même, elles restent bien inférieures à la pression sous-glottique en phonation qui se situe aux environs de 7 000 Pa. Pour enregistrer un son, il est donc nécessaire d'utiliser des transducteurs adaptés à cet ordre de grandeur.

Le microphone, capteur de son par excellence, permet d'enregistrer les variations de pression de l'air se situant dans la gamme précédemment décrite. Son rôle est de fournir une tension électrique proportionnelle aux variations de pression acoustique. Il existe différents types de microphones. Les plus répandus utilisent une technologie électrodynamique. Simples d'utilisation, économiques, ils permettent de faire des enregistrements de qualité moyenne. Les microphones à condensateur ou à électret sont des dispositifs de très bonne qualité mais de prix élevé. De plus, ils nécessitent une alimentation à piles ou par des circuits dits «fantômes». Cependant, ils restent idéals pour des enregistrements professionnels de haute fidélité.

II. Les grandeurs physiques et perceptives d'un son pur

Un son pur est constitué d'une oscillation simple vibrant à une période et une amplitude données (Fig. 1). Un tel son est appelé aussi onde sinusoïdale ou signal élémentaire. La tonalité du téléphone en est un bon exemple.

1. Période, fréquence, pulsation, hauteur

a. Définitions des grandeurs physiques liées à la fréquence

La **période** T correspond à la durée d'un cycle de vibration (Fig. 1). Elle s'exprime en secondes (s) ou en millisecondes (ms). La **fréquence** F est l'inverse de la période et s'exprime en hertz (Hz). Elle correspond au nombre de cycles par seconde.

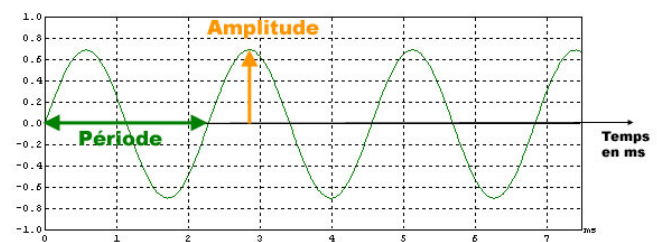


Fig. 1. Décours temporel d'un son pur avec ses caractéristiques propres (période et amplitude).

$$F = 1/T$$

Parfois, est mentionnée la notion de **pulsation** (ω) qui est directement liée à la fréquence par la relation $\omega = 2\pi \cdot F$ et s'exprime en rad/s.

Fréquence (Hz)	Période (s)	Période (ms)	
10	0.1	100	Grave ⇔ Basses Fréquences
50	0.02	20	
100	0.01	10	
200	0.005	5	Aigu ⇔ Hautes Fréquences
500	0.002	2	
1000	0.001	1	
2000	0.0005	0.5	

Tableau 1. Quelques équivalences entre fréquence et période (on rappelle : 1 s = 1000 ms).

La figure 2 permet d'illustrer visuellement la relation entre un son grave ⇔ basse fréquence ⇔ longue période et un son aigu ⇔ haute fréquence ⇔ courte période. Une façon de calculer manuellement la fréquence d'un son à partir de la visualisation du

signal (oscillogramme) consiste à compter le nombre de cycles relevés pour un temps donné et à diviser ce nombre de cycles par la durée en faisant attention aux unités (1 ms = 0.001 s).

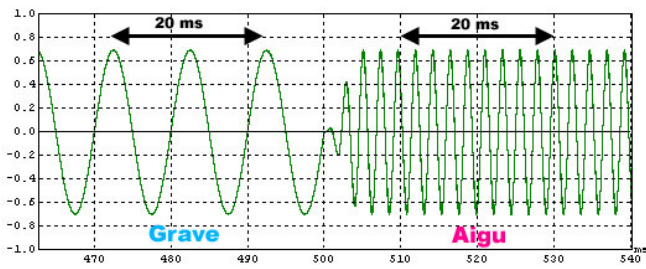


Fig. 2. Décours temporel de sons purs à 2 fréquences différentes.

Dans la 1^{re} partie, on mesure 2 cycles pour 20 ms (= 0.02 s) $\Leftrightarrow F = 2/0.02 = 100$ Hz. Puis, dans la 2^e partie, 9 cycles pour 20 ms $\Leftrightarrow F = 9/0.02 = 450$ Hz.

b. Perception de la fréquence, notions d'intervalle, demi-tons et octave

La hauteur est une notion liée essentiellement à la perception de la fréquence d'un son, bien que le niveau sonore, le timbre et la durée du signal puissent aussi influencer la perception de la hauteur (Kitantou, 1987). Le système auditif humain est sensible aux vibrations comprises entre 20 Hz et 20 000 Hz. Les vibrations en dessous de 20 Hz sont appelées infrasons et peuvent être perçues par la paroi abdominale, propriété exploitée dans certaines salles de concert ou de cinéma. Les ultrasons se situent eux au-dessus de 20 000 Hz. Certains animaux, comme les chiens, les perçoivent très bien. Ces valeurs restent générales et chaque individu a des spécificités. L'âge contribue notamment à réduire cette plage de fréquences perceptibles.

Il est important de noter que la sensation relative de hauteur est logarithmique. En effet, d'un point de vue perceptif, l'écart de 130 Hz entre un do3 (262 Hz) et un sol3 (392 Hz) est perçu de façon équivalente à celui entre un do4 (523 Hz) et un sol4 (784 Hz) alors que dans ce cas, l'écart est de 261 Hz. En fait, c'est le rapport, appelé intervalle entre les deux notes, qui est constant et qui vaut quasiment 1.5 dans les deux cas. Par définition, l'intervalle entre deux hauteurs ne se mesure pas par la différence des fréquences mais par leur rapport. C'est pour rendre compte de cette propriété que les notions de demi-tons et d'octave ont été introduites. L'octave correspond à un doublement de fréquence. Ainsi, 110 Hz, 220 Hz, 440 Hz, 880 Hz sont des fréquences séparées respectivement d'une octave.

Nous ne rentrerons pas dans l'épineux problème de la constitution de la gamme diatonique, qui fait intervenir des intervalles non constants, et qui ont la propriété de s'exprimer par des fraction simples, plus agréables à l'oreille, mais qui posent des problèmes de transposition. Nous resterons dans le cadre de la gamme chromatique tempérée, uniquement fondée sur des critères physiques. Dans cette gamme utilisée dans la musique occidentale, une octave a été divisée en 12 intervalles égaux appelés demi-tons qui donnent lieu aux douze notes de la gamme (do, do#, ré, ré#, mi, fa, fa#, sol, sol#, la, la#, si). D'un point de vue mathématique (Kitantou, 1987), la valeur des demi-tons successifs est régie par une suite géométrique de raison $\alpha = 1.06$, c'est-à-dire que pour obtenir le demi-ton suivant une fréquence f , il suffit de multiplier f par 1.06.

En effet, sachant que dans la gamme chromatique, l'intervalle α est constant, que pour augmenter d'un demi-ton, on multiplie la fréquence par α , on peut écrire : $f_1 = \alpha \cdot f_0$; $f_2 = \alpha \cdot f_1 = \alpha^2 \cdot f_0$;
 $f_3 = \alpha \cdot f_2 = \alpha^3 \cdot f_0 \dots$, $f_{12} = \alpha \cdot f_{11} = \alpha^{12} \cdot f_0$
 Sachant que cette augmentation renouvelée 12 fois donne lieu à une octave qui correspond à un doublement de fréquence, on peut écrire :
 $f_{12} = 2 \cdot f_0$. Les deux égalités donnent donc $\alpha^{12} \cdot f_0 = 2 \cdot f_0 \Leftrightarrow$
 $\alpha^{12} = 2 \Leftrightarrow \alpha = 2^{1/12} \Leftrightarrow \alpha = 1,05946$.

De la même façon, on peut montrer que le nombre de demi-tons séparant une fréquence f d'une autre fréquence f_0 est le suivant :

$$N \text{ (en } \frac{1}{2} \text{ tons)} \approx 40 \cdot \text{Log}_{10} (f/f_0)$$

En effet, appelons α l'intervalle entre deux demi-tons. Nous cherchons l'intervalle N en demi-tons entre f_0 et f . Cela se traduit par l'équation :
 $\alpha^N = f / f_0$. Nous avons vu précédemment que $\alpha = 2^{1/12}$, ce qui donne alors $2^{N/12} = f / f_0$. En appliquant un logarithme décimal de chaque côté de l'égalité, on obtient $\text{Log}_{10} (2^{N/12}) = \text{Log}_{10} (f / f_0) \Leftrightarrow$
 $(N/12) \cdot \text{Log}_{10} (2) = \text{Log}_{10} (f / f_0) \Leftrightarrow N \cdot [\text{Log}_{10} (2) / 12] = \text{Log}_{10} (f / f_0)$
 $\Leftrightarrow N = [12 / \text{Log}_{10} (2)] \cdot \text{Log}_{10} (f / f_0)$
 Sachant que $12 / \text{Log}_{10} (2) = 39,8631$, on obtient la relation
 $N \text{ (en } \frac{1}{2} \text{ tons)} \approx 40 \cdot \text{Log}_{10} (f / f_0)$.

Cette relation est intéressante à retenir car elle est fréquemment utilisée pour connaître, par exemple, l'étendue vocale d'un chanteur ou plus généralement d'un locuteur.

2. Amplitude, énergie, intensité et puissance sonore

a. Définitions des grandeurs physiques liées au niveau sonore

L'amplitude correspond à l'ampleur de la vibration (Fig. 1) et s'exprime en pascal (Pa) car il s'agit d'une mesure de pression acoustique. Elle peut aussi s'exprimer en newton/m² car une pression correspond à une force (N) s'exerçant sur une surface (m²). Sachant qu'un auditeur jeune est capable de percevoir une pression acoustique de l'ordre de 20 µPa (0.00002 Pa), on peut montrer que la force exercée sur le tympan est équivalente à la force de pesanteur qui s'exercerait sur une masse d'environ 0,16 µg, c'est-à-dire un cube d'eau de 5,5 centièmes de mm de côté. Autant dire que nous nous situons dans le domaine de la micro-mécanique. D'un point de vue perceptif, plus un son a une amplitude importante, plus il sera perçu fort.

Une source sonore génère de l'énergie mécanique qui est transportée, via la vibration acoustique, sous forme cinétique (vitesse des molécules d'air) et potentielle (compression, détente). Cette énergie s'exprime en joules (J) ou encore en newton.m car une énergie est équivalente à une force (N) appliquée sur une distance (m).

La puissance d'une source sonore correspond à son débit d'énergie, c'est-à-dire qu'elle rend compte de la quantité d'énergie acoustique délivrée par seconde et se mesure en joules/s, c'est-à-dire en watts (1 W = 1 J/s). L'intensité acoustique est, en un point donné de l'espace, le flux moyen de la puissance par unité de surface. Elle s'exprime donc en W/m². Liénard et François (1983) montrent que cette intensité sonore, parfois appelée niveau sonore, est liée au carré de la pression acoustique dans le cas d'une onde progressive plane.

Pour résumer, on peut lier les différentes grandeurs physiques de la façon suivante :

- Une source sonore diffuse de l'énergie mécanique notée E (en joules) sous forme d'énergie cinétique et potentielle.
- Une source sonore qui diffuse une énergie acoustique E en un temps t possède une puissance W (en watts) = E/t.
- Considérons une surface S. Supposons que pendant un temps t, elle soit traversée par une énergie E. L'intensité sur cette surface sera égale à : $I = E / (t \cdot S) = W / S$ (watt/m²).
- De façon littérale (Liénard et François, 1983), $I = A_{eff}^2 / \rho c$ où I est l'intensité en watt/m², A_{eff} est la pression efficace, ρ est le densité du milieu (kg/m³) et c la vitesse de propagation du son (m/s). Dans l'air à 20°C, ρc = 400 MKS.

b. La notion de décibel

Sachant que la puissance sonore peut varier d'une fraction de microwatt (1 µW = 0.000001 W) pour le tic-tac d'une montre à des dizaines de kilowatts (1 kW = 1000 W) pour un réacteur d'avion, il s'avère peu pratique de manipuler une telle étendue de valeurs (Bourcet et Liénard, 1987). En effet, le rapport entre les plus fortes et les plus faibles mesures de puissance acoustique est de l'ordre du milliard. Aussi, afin de réduire cette très large gamme de valeurs, les physiciens ont pris l'habitude d'exprimer la puissance et l'intensité en décibels (dB Intensity Level ou dB IL). Par définition, un bel est le logarithme décimal d'un rapport de puissances ou d'intensités : Log₁₀ (I1/I2). Le terme a été choisi en hommage à Graham Bell, pionnier et fondateur du laboratoire des télécommunications. Par extension, le décibel est le dixième de bel et sa définition est donc 10.Log₁₀ (I1/I2). Appliquée à une mesure acoustique, la relation devient la suivante :

$$dB \text{ IL} = 10 \cdot \text{Log}_{10} (I/I_0), \text{ où } I \text{ est l'intensité en } W/m^2 \text{ et } I_0 \text{ est un niveau de référence fixé à } 10^{-12} W/m^2$$

Comme l'intensité acoustique est proportionnelle au carré de la pression acoustique (cf. §a.), il est tentant de sortir les carrés du logarithme pour arriver à la définition du décibel en pression (dB Sound Pressure Level ou dB SPL). En effet, on rappelle que Log(Kⁿ) = n.Log(K). A partir de la définition précédente, on obtient les égalités suivantes :

$$dB = 10 \cdot \text{Log}_{10} [I/I_0] = 10 \cdot \text{Log}_{10} [(A^2/400)/(A_0^2/400)] \\ = 10 \cdot \text{Log}_{10} [A^2/A_0^2] = 10 \cdot \text{Log}_{10} [A/A_0]^2 = 2 \cdot 10 \cdot \text{Log}_{10} [A/A_0] \\ = 20 \cdot \text{Log}_{10} [A/A_0], \text{ d'où le résultat :}$$

$$dB \text{ SPL} = 20 \cdot \text{Log}_{10} (A/A_0), \text{ où } A \text{ est la pression en Pa} \\ \text{et } A_0 \text{ est le niveau de référence de :} \\ 0.00002 \text{ Pa (seuil de l'audition).}$$

Avec de telles transformations mathématiques, il est facile de montrer les relations suivantes.

De même, nous présentons ci-après les équivalences

Amplitude	Intensité ou Puissance	dB
Multipliée par 1.41	Double	+3 dB
Double	Quadruple	+6 dB
Multipliée par 3.16	Décuple	+10 dB
Décuple	Centuple	+20 dB

Amplitude	Intensité ou Puissance	dB
Divisée par 1.41	Divisée par 2	-3dB
Divisée par 2	Divisée par 4	- 6dB
Divisée par 3.16	Divisée par 10	-10 dB
Divisée par 10	Divisée par 100	-20 dB

Tableau 2. Mesures en dB des modifications d'amplitude, d'intensité ou de puissance sonore.

entre pressions et dB SPL, ceci en relation avec différents environnements sonores donnés en exemple par Durrant et Lovrinic (1984).

Il est important de signaler que si cette

200 Pa = 140 dB SPL (avion à réaction)

20 Pa = 120 dB SPL (concert rock) seuil de douleur

2 Pa = 100 dB SPL (métro)

0.2 Pa = 80 dB SPL (musique forte d'un poste de radio)

0.02 Pa = 60 dB SPL (conversation normale)

0.002 Pa = 40 dB SPL (local calme)

0.0002 Pa = 20 dB SPL (bruissement de feuilles)

0.00002 Pa = 0 dB SPL (seuil de l'audition)

Tableau 3. Mesures d'amplitude en Pa et dB pour différents environnements sonores.

transformation en décibels s'est généralisée en acoustique, cela provient de la bonne corrélation entre cette échelle logarithmique et la perception auditive. En effet, d'après la loi de Weber-Fechner, qui s'applique d'ailleurs à d'autres sens que l'audition, la sensation varie comme le logarithme de l'excitation (Kitantou, 1987). Ceci a d'importantes conséquences pratiques. Un affaiblissement de trois décibels correspond à une division par deux de la puissance sonore, mais cette baisse est à peine sensible. Si le niveau vocal d'un locuteur baisse de trois décibels, l'auditeur s'en apercevra pendant la seconde qui suit l'atténuation, mais si le locuteur s'en va et revient le lendemain avec une baisse de trois ou même de cinq décibels, l'auditeur ne s'en apercevra pas. Or, diminuer de six décibels, c'est déjà diviser par quatre la puissance sonore (Bourcet et Liénard, 1987).

c. Les notions d'intensité subjective, de courbes isotoniques et de pondérations soniques

Nous avons vu que la sensation auditive du niveau sonore n'est pas directement liée à la grandeur physique de l'amplitude mais qu'elle varie comme le logarithme de l'excitation sonore. De plus, Fletcher et Munson (1933) ont montré qu'une autre non-linéarité intervient entre perception et réalité physique, liée aux fréquences des sons. Ils constataient, par exemple, qu'un son de 40 dB SPL à 1000 Hz devait être produit à 60 dB SPL à 100 Hz pour être perçu au même niveau sonore. Ces travaux ont donné lieu, par la suite, à différentes échelles perceptives comme l'échelle des phones, les courbes d'isotonie (Fig. 3).

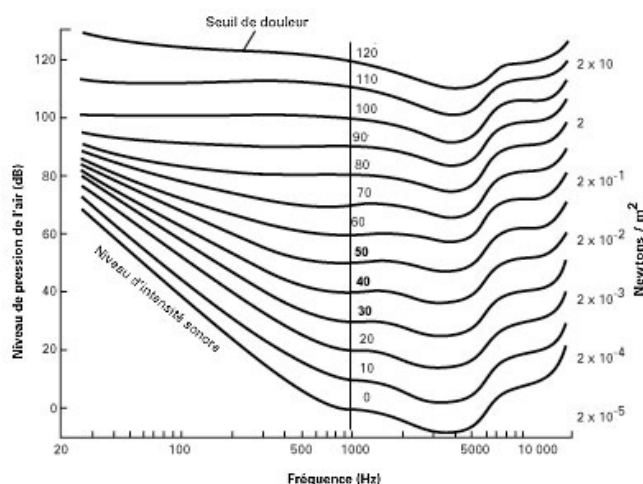


Fig. 3. Courbes de niveaux d'isotonie.

La référence étant fixée à 1000 Hz, les points appartenant à une même courbe d'isotonie correspondent à une intensité perceptive équivalente. Par exemple, 40 dB à 1000 Hz \Leftrightarrow 60 dB à 100 Hz.

Une norme appelée IEC 179 a finalement été proposée pour permettre de transposer l'intensité physique mesurée vers une intensité dite «subjective». Ont été ainsi créées des courbes de pondération avec lesquelles le niveau SPL est plus ou moins atténué dans chaque bande d'octave pour tenir compte des non-linéarités observées de la perception auditive en fonction de la fréquence. Les pondérations A, B et C (Fig. 4) sont fondées respectivement pour les sons de niveau faible (40 phones \Leftrightarrow 40 dB SPL à 1000 Hz), moyen (70 phones \Leftrightarrow 70 dB SPL à 1000 Hz) et fort (100 phones).

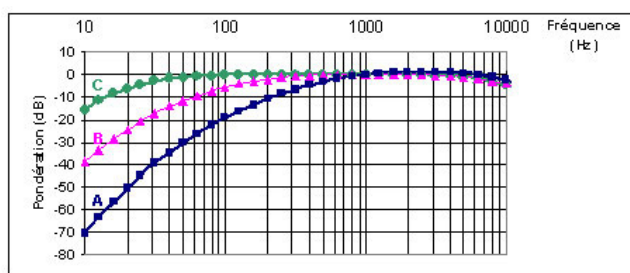


Fig. 4. Courbes de pondération sonique A, B, C permettant d'obtenir une intensité subjective à partir de la mesure d'une intensité SPL (norme IEC 179, table des données dans la notice technique de Brüel et Kjaer, 1967).

Le sonomètre est l'instrument privilégié pour la mesure des intensités sonores. La plupart de ces appareils sont équipés de «filtres» permettant de mesurer des dB SPL mais aussi des dB(A), échelle généralement utilisée dans les mesures de nuisances sonores. Le dB(A) a aussi été proposé par l'Union Européenne des Phoniatres pour la phonétographie (Schutte et Seidner, 1983).

d. Les phénomènes de propagation sonore

Pour compléter ces notions d'amplitude et d'intensité sonore, il faut exposer quelques concepts liés à la propagation des sons dans l'atmosphère. En effet, en se propageant dans l'air, l'onde sonore subit un certain nombre de phénomènes tels que l'absorption, la réflexion, la réverbération... Comme l'ont détaillée Bourcet et Liénard (1987), l'atténuation d'une onde sonore dans l'atmosphère est provoquée par divers phénomènes : pertes mécaniques liées à la viscosité de l'air, échanges thermiques mais aussi turbulences et caractères aléatoires de l'atmosphère. Toutefois, ces manifestations n'interviennent que pour de longues distances (plusieurs centaines de mètres) en extérieur, ce qui nous éloigne de nos préoccupations. Nous n'aborderons que la notion d'atténuation géométrique, appelée aussi dilution acoustique.

Si l'on se place dans le cas d'une source sonore qui diffuse de façon sphérique dans l'espace, l'énergie émise va se diluer sur des sphères de plus en plus grandes, dont la surface varie comme le carré de la distance entre la source et le point de réception. L'intensité en W/m^2 sera ainsi d'autant plus faible que l'auditeur ou l'appareil de mesure est à une distance importante de la source ; de façon exacte, l'intensité diminue comme l'inverse du carré de la distance, équivalent à une loi en $20 \cdot \text{Log}_{10}(d)$. Autrement dit, le niveau diminue de 6 dB quand on double la distance à la source. Par conséquent, en toute

rigueur, quand on fournit une valeur d'amplitude ou d'intensité acoustique, il faut systématiquement préciser à quelle distance de la source la mesure a été prise. Par convention, la distance de référence utilisée pour les mesures en acoustique est de 1 mètre. Toutefois, dans des environnements bruités, il est préférable de se placer à 30 cm. Une chose est sûre, dans les études comparatives d'intensité, il est nécessaire que la même distance soit respectée entre la source sonore (ex. : un locuteur) et le microphone ou autre dispositif d'enregistrement.

III. L'onde sonore complexe et l'analyse spectrale

1. Signal réel et signaux élémentaires

La notion de son pur, précédemment exposée, reste une illustration pédagogique mais ne se rencontre que très peu dans la nature. Dans la réalité, un son est complexe et formé d'une superposition de vibrations simples ayant des amplitudes et des fréquences différentes. Le théorème de Fourier démontre qu'un signal de forme quelconque peut être décomposé en un ensemble de signaux élémentaires d'amplitude et de fréquences déterminées. Inversement, un mélange de signaux sinusoïdaux émis simultanément produit un signal complexe (Fig. 5).

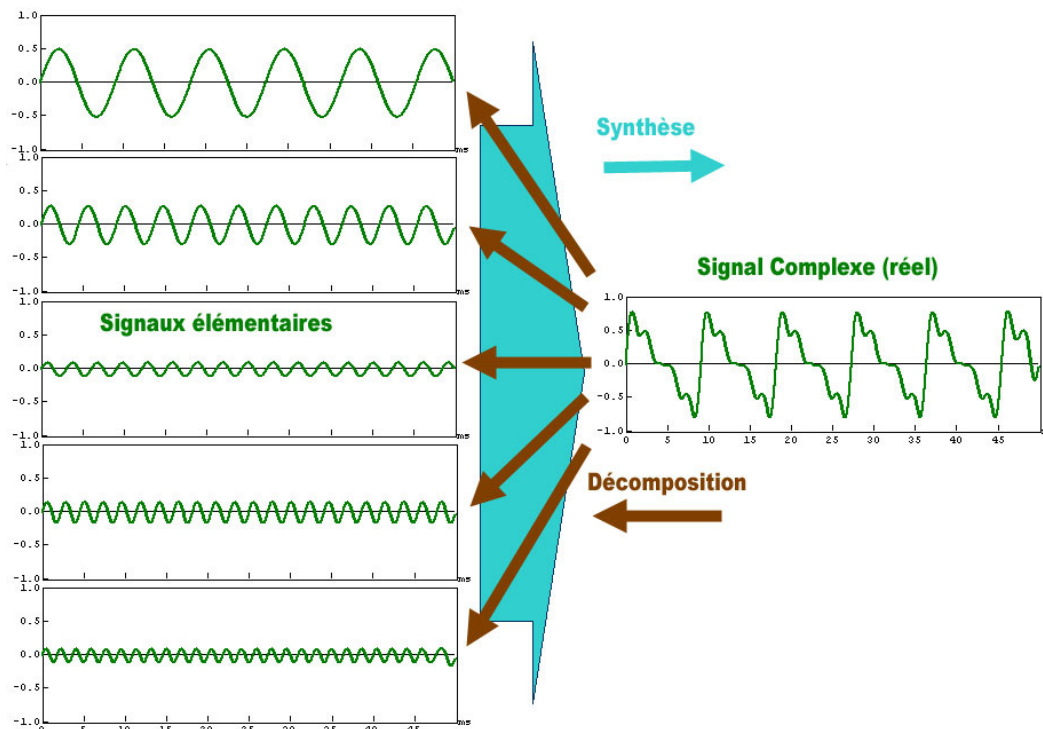


Fig. 5 : signaux élémentaires (à gauche) et signal complexe (à droite).

(a) synthèse : la superposition de vibrations simples génère un son complexe (de la gauche vers la droite)

(b) décomposition : une onde complexe peut être décomposée en un ensemble de vibrations simples (de droite à gauche).

2. Le spectre

a. Définition du spectre

Le spectre d'un signal, obtenu par Transformée de Fourier, indique les caractéristiques en amplitude et en fréquence des ondes sinusoïdales élémentaires qui composent le signal (Fig. 6).

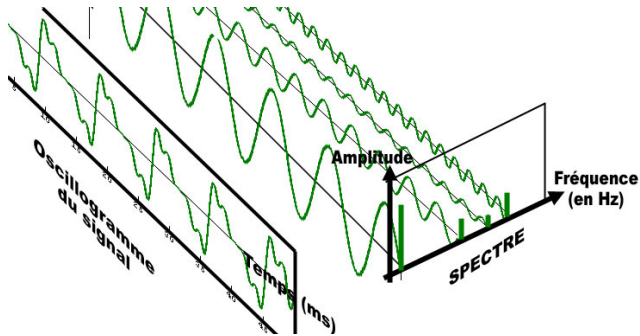


Fig. 6. Représentation du signal dans le domaine temporel (oscillogramme) et dans le domaine des fréquences (spectre).

b. Sons périodiques *versus* sons apériodiques

Un son est dit périodique s'il oscille de façon régulière (cyclique). Le cycle de base correspond à la fréquence fondamentale notée F_0 . Les fréquences multiples de la fondamentale sont appelées les harmoniques. On distingue parfois les harmoniques paires ($2 F_0$, $4 F_0$, ...) et impaires ($3 F_0$, $5 F_0$, ...). Le spectre d'un signal périodique comporte des raies (Fig. 7) correspondant à la fréquence fondamentale et ses harmoniques (ex. : voyelles, consonnes voisées, flûte, piano...).

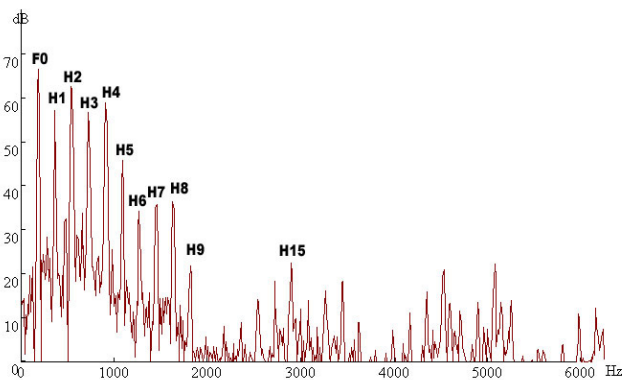


Fig. 7. Spectre d'une voyelle /a/.

La 1^{re} raie correspond à la F_0 , la 2^e raie à la 1^{re} harmonique ($=2F_0$), la 3^e raie à la 2^e harmonique ($=3F_0$)...

Tous les sons ne sont pas périodiques. Un coup de cymbale, le bruit du vent, les consonnes /p/, /t/, /k/, /f/, /s/... sont des sons apériodiques. Cela signifie qu'on ne retrouve pas dans le déroulement temporel du signal un aspect cyclique. De même, dans le spectre, la structure de raie a disparu au bénéfice d'une répartition diffuse (Fig. 8).

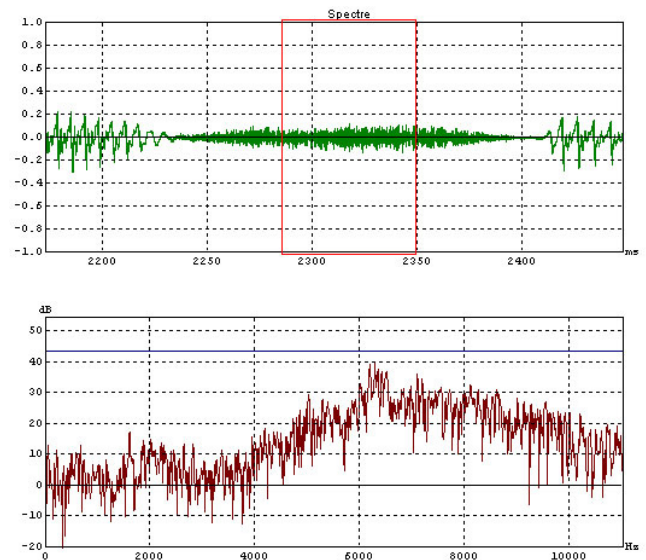


Fig. 8. Oscillogramme (en haut) et spectre (en bas) de la consonne /s/ dans le mot «façon».

Le déroulement temporel a une structure non périodique (bruit de friction) et le spectre n'a pas de structure de raie car ce phonème est non voisé.

c. Le timbre

Le timbre correspond à la sensation auditive du spectre. A ce terme sont souvent associées les notions de signature, couleur, identité du son. En effet, c'est par la perception précise du timbre qu'on reconnaît à l'oreille les différents sons de la parole, les locuteurs, les instruments de musique... On parle souvent de timbre clair quand la composition spectrale du son contient des hautes fréquences. Par contre, un timbre sombre est souvent lié à un son essentiellement constitué de basses fréquences.

3. L'amplitude d'un son complexe : valeur efficace, intensité RMS

Il est délicat de parler d'amplitude instantanée pour un signal complexe car il est composé de signaux élémentaires à amplitudes et fréquences différentes. Aussi a été introduite la notion d'amplitude efficace appelée habituellement «intensité RMS» (Root Mean Square), qui correspond approximativement à l'amplitude moyenne de l'ensemble des signaux élémentaires qui le composent. Cette grandeur, appelée abusivement intensité alors qu'elle correspond plutôt à une amplitude, se calcule en intégrant le signal temporel sur une durée de l'ordre du centième de seconde et s'exprime en dB (Fig. 9). Le terme Root Mean Square provient du fait que le calcul consiste à prendre la racine carrée (root) de la somme des carrés (square) des valeurs du signal sur une durée moyenne.

IV. Les aspects dynamiques de l'onde sonore

Que ce soit en fréquence, en amplitude ou en composition spectrale, les sons ont des caractéristiques qui évoluent dans le temps.

1. Les variations de fréquence fondamentale

Que ce soit pour de la musique ou de la parole, les variations mélodiques sont liées aux variations de la fréquence fondamentale F_0 . Celle-ci varie d'un individu à l'autre, la F_0 des hommes étant généralement plus basse que celle des femmes mais pour un même individu, elle varie aussi au cours de son élocution, donnant lieu à la musicalité de la communication chantée et parlée. Il faut signaler que pour la parole, la courbe de F_0 est fréquemment interrompue durant des pauses silencieuses ou au cours de la production de consonnes non voisées comme /p/, /t/, /k/, /f/, /s/ qui ont la particularité d'être des sons aperiodiques (Fig. 9).

Les fonctions linguistiques liées à la fréquence fondamentale sont nombreuses et nous n'entrerons pas dans les détails dans ce chapitre. Toutefois, il faut signaler que cette capacité à moduler la fréquence fondamentale du signal de parole rend compte d'une bonne maîtrise vocale, les patients dysphoniques ayant pour leur part une étendue réduite.

2. La dynamique d'amplitude

Que ce soit pour de la musique ou de la parole, l'amplitude efficace (cf. § III.3) de l'onde sonore varie dans le temps. C'est par ce phénomène que le musicien module le niveau sonore de son instrument en jouant pianissimo, piano, forte ou fortissimo. De la même façon, l'onde sonore, vecteur de la communication parlée est caractérisée par des variations d'intensité liées à des phénomènes intrinsèques. Ainsi, globalement, les voyelles sont plus énergétiques que les consonnes (Fig. 9). D'autre part, le locuteur peut utiliser cette fonctionnalité de façon volontaire pour effectuer des contrastes comme de la focalisation ou de l'insistance. Enfin, il faut signaler que cette capacité à moduler en amplitude le signal vocal rend compte d'une bonne maîtrise vocale, les patients hypophoniques ayant pour leur part une étendue réduite.

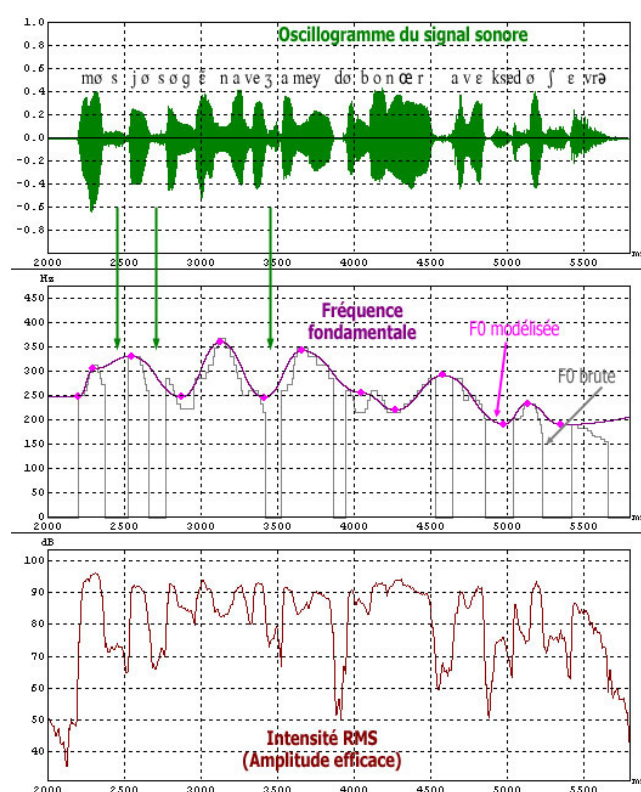


Fig. 9. Oscillogramme du signal de parole, fréquence fondamentale et amplitude efficace sur l'extrait de parole «Monsieur Seguin n'avait jamais eu de bonheur avec ses deux chèvres». Les flèches vertes indiquent les parties non voisées (\Leftrightarrow pas de F_0 brute).

3. Les variations spectrales : la spectrographie

Un son qui n'a pas de variations spectrales est équivalent à une image monochrome. Bien évidemment, la plupart des ondes sonores ont une composition spectrale qui varie dans le temps, en particulier la parole. Les spectres présentés aux figures 7 et 8 n'analysent qu'une partie du signal : une voyelle /a/ sur la figure 7, une consonne /s/ sur la figure 8. Un spectrogramme est une représentation temps-fréquence qui permet de visualiser l'évolution du spectre au cours du temps. Il faut signaler que le terme sonographe ou même «sona» est entré dans l'usage mais il s'agit d'une marque commerciale de spectrographe.

Le principe est de sélectionner une partie du signal sur laquelle est calculé un spectre (Fig. 10a). L'application de seuils permet de séparer les niveaux et d'attribuer une couleur par «tranche de niveau» (Fig.10b). En général, on utilise des nuances de gris, le gris clair étant réservé aux faibles niveaux, le noir révélant les pics d'énergie spectrale maximale. Puis, on effectue une rotation du spectre qui déplace l'axe

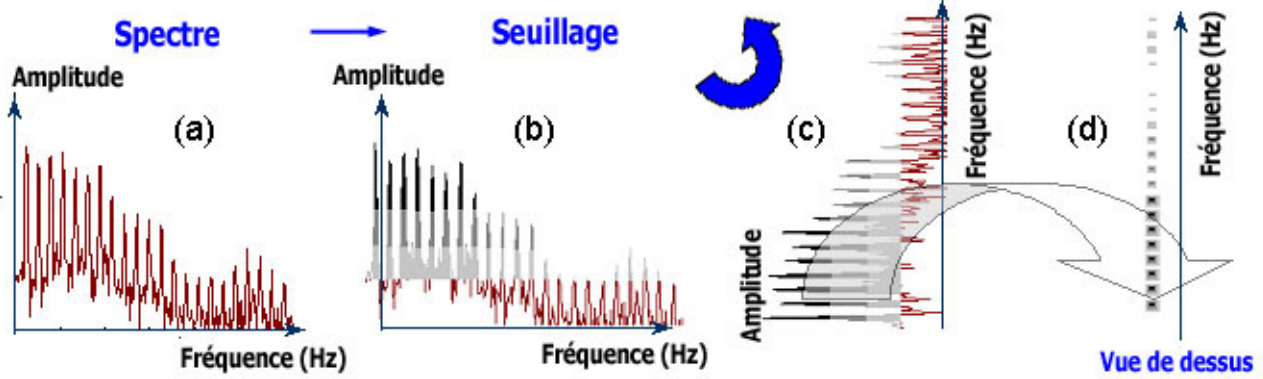


Fig. 10. Principe de la construction spectrographique : (a) spectre, (b) seuillage, (c) rotation des axes, (d) renversement et vue de dessus.

des fréquences verticalement (Fig.10c). Le spectre est finalement «redressé» et vu par dessus (Figure 10d) : le niveau d'amplitude est lié au niveau de couleur. L'opération est renouvelée pour le segment temporel suivant et la nouvelle représentation spectrale est concaténée à la précédente jusqu'à fournir une vaste matrice temps-fréquence-amplitude (Fig. 11). On peut comparer une telle représentation à une carte de géographie physique où les faibles énergies sont assimilées aux plaines, les énergies moyennes aux plateaux ou collines et les fortes énergies aux sommets montagneux. L'axe des fréquences est l'axe sud-nord, le déroulement temporel est l'axe ouest-est.

Le spectrogramme étant une concaténation de spectres effectués sur des portions juxtaposées de signal, le choix de la taille de la fenêtre d'analyse délimitant chaque portion a une importance. En effet, plus la taille de la fenêtre temporelle sera importante, plus le spectre aura une précision fréquentielle importante. Par contre, la précision temporelle sera médiocre du fait de l'empan temporel imposant de la fenêtre.

Sur la figure 11, on remarque que le réglage en «bandes larges» permet une bonne définition temporelle car les fenêtres d'analyse sont courtes. Il permet ainsi de repérer précisément les événements phonétiques (ex. : explosion du /t/, frontières phonémiques...).

Par construction, si la fenêtre temporelle comporte N échantillons de signal, le spectre comportera $N/2$ points de mesures spectrales. Appelons B la bande passante du signal (qui correspond à la fréquence maximale observable), la résolution spectrale sera alors de $R = B/(N/2) = 2.B / N = F_{ech} / N$ où F_{ech} est la fréquence d'échantillonnage, qui est le double de la bande passante (théorème de Fourier). Il faut signaler que dans les algorithmes de Transformée de Fourier Rapide (Fast Fourier Transform ou FFT), N est choisi par construction comme une puissance de 2 (ex : $N = 64, 128, 256, 1024, 2048, 4096$ points...).

Prenons par exemple un fichier échantillonné à 22050 Hz, une fenêtre de 256 points ($\Leftrightarrow 11.6$ ms) permettra d'obtenir une résolution fréquentielle de 86.13 Hz, c'est-à-dire qu'on obtiendra une valeur du spectre pour :

$$F = 1*86.13, 2*86.13 = 172, 3*86.13 = 258, 4*86.13 = 344, \dots, 10*86.13 = 861, \dots, 128*86.13 = 11025$$

Une telle résolution, appelée «bandes larges», est insuffisante pour faire apparaître les raies de F_0 et des harmoniques (Fig. 11b). Par contre, pour le même signal, une fenêtre de 2048 points ($\Leftrightarrow 93$ ms) permet d'obtenir une résolution fréquentielle de 10.76 Hz, qui cette fois est suffisante pour observer des spectres de raies, d'où sa dénomination de «bande étroite» (Fig. 11c).

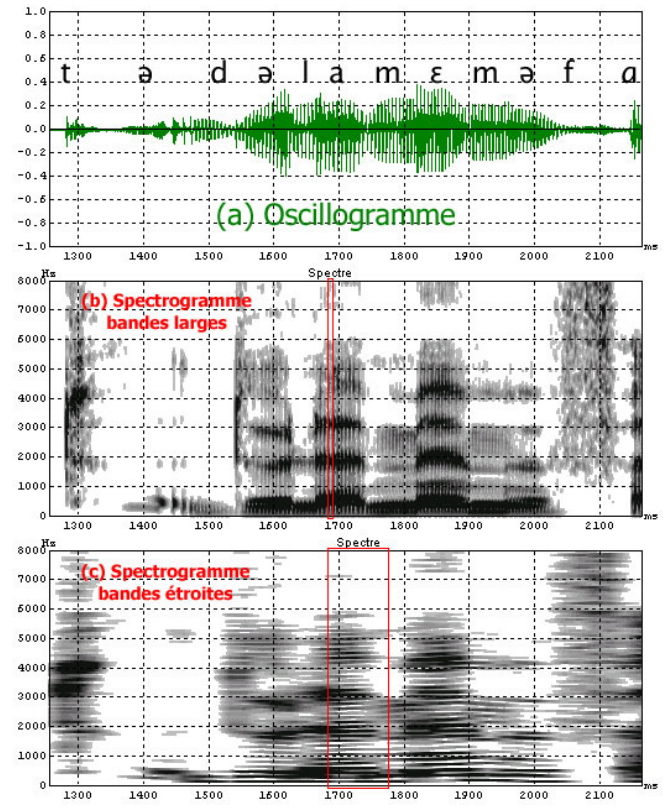


Fig. 11. Oscillogramme et spectrogrammes de l'extrait «il perdait tout TES DE LA MEME FAÇON».

Inversement, avec un réglage en bandes étroites, les frontières deviennent floues mais cela permet, par contre, une bonne définition fréquentielle pour la mise en évidence de la fréquence fondamentale et de ses harmoniques. La taille de la fenêtre d'analyse est visible sur chacun des spectrogrammes sous forme de rectangle rouge étroit pour le spectrogramme en bandes larges, étendu pour le spectrogramme en bandes étroites.

V. Conclusion

L'analyse acoustique de la parole, notamment pathologique, ne nécessite pas un savoir encyclopédique en mécanique ou sur les théories des ondes. Quelques connaissances fondamentales sont nécessaires pour appréhender les phénomènes observés. Par contre, l'expertise se construit surtout par la multiplication et la confrontation des observations du signal acoustique. Il ne faut pas non plus perdre de vue le décalage qui existe entre les données physiques observables et la perception, leurs relations étant très souvent non linéaires comme la sensation logarithmique de l'amplitude, comme la sensation relative de hauteur ou encore les phénomènes mixtes comme les courbes d'isotonie. Nous ne mentionnons là que des phénomènes de bas niveau. Que dire alors de la perception de la qualité vocale qui fait intervenir un grand nombre de facteurs ? On peut penser que c'est par l'utilisation conjointe de méthodes (analyse acoustique et jugement auditif) que l'expérimentateur pourra obtenir des réponses à son questionnement.

Références

1. Botte MC, Canévet G, Demany L, Sorin C. (1988). *Psychoacoustique et perception auditive. Série Audition. INSERM/SFA/CENT.*
2. Bourcet P, Liénard P. *Acoustique fondamentale. In Mercier D (éd.). Le livre des techniques du son, vol. 1 : Notions fondamentales. Editions Eyrolles, Fréquences. Paris. 1987, pp 13-43.*
3. Durrant JD, Lovrinic JH. (1984). *Bases of Hearing Science. 2nd Edition. Williams et Wilkins. Baltimore.*
4. Fletcher HF, Munson WA. *Loudness, its definition, measurement and calculation. J Acoust Soc Am 1933, 5: 82-108.*

5. Kitantou M. *La perception auditive. In Mercier D (éd.). Le livre des techniques du son, vol. 1 : Notions fondamentales. Editions Eyrolles, Fréquences. Paris. 1987, pp 155-181.*
6. Liénard P, François P (1983). *Acoustique industrielle et environnement. Tome 1: acoustique physique et perceptive. Collection de la Direction des Etudes et Recherches d'E.D.F. monographie publiée sous les auspices du Groupement des Acousticiens de Langue Française. Editions Eyrolles. Paris.*
7. Schutte HK, Seidner W. *Recommendations by the Union of European Phoniaticians (UEP) : Standardizing Voice Area measurements/ Phonetography. Folia Phoniatic 1983, 35: 286-288.*